Mixed Integer Programming Approaches for Experimental Design

Juan Pablo Vielma
Massachusetts Institute of Technology

Joint work with Denis Saure

DRO brown bag lunch seminars, Columbia Business School
## Motivation: (Custom) Product Recommendations

<table>
<thead>
<tr>
<th>Feature</th>
<th>SX530</th>
<th>RX100</th>
</tr>
</thead>
<tbody>
<tr>
<td>Zoom</td>
<td>50x</td>
<td>3.6x</td>
</tr>
<tr>
<td>Prize</td>
<td>$249.99</td>
<td>$399.99</td>
</tr>
<tr>
<td>Weight</td>
<td>15.68 ounces</td>
<td>7.5 ounces</td>
</tr>
<tr>
<td>Prefer</td>
<td>✔️</td>
<td>☐</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Feature</th>
<th>TG-4</th>
<th>G9</th>
</tr>
</thead>
<tbody>
<tr>
<td>Waterproof</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Prize</td>
<td>$249.99</td>
<td>$399.99</td>
</tr>
<tr>
<td>Weight</td>
<td>7.36 lb</td>
<td>7.5 lb</td>
</tr>
<tr>
<td>Prefer</td>
<td>☐</td>
<td>✔️</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Feature</th>
<th>TG-4</th>
<th>Galaxy 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Waterproof</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Prize</td>
<td>$249.99</td>
<td>$399.99</td>
</tr>
<tr>
<td>Viewfinder</td>
<td>Electronic</td>
<td>Optical</td>
</tr>
<tr>
<td>Prefer</td>
<td>✔️</td>
<td>☐</td>
</tr>
</tbody>
</table>

We recommend:
Motivation: (Custom) Product Recommendations

- Toubia, Hauser and Dahan (2003)
“Towards” Optimal Product Recommendation

• Find enough information about preferences to recommend

<table>
<thead>
<tr>
<th>Feature</th>
<th>SX530</th>
<th>RX100</th>
</tr>
</thead>
<tbody>
<tr>
<td>Zoom</td>
<td>50x</td>
<td>3.6x</td>
</tr>
<tr>
<td>Prize</td>
<td>$249.99</td>
<td>$399.99</td>
</tr>
<tr>
<td>Weight</td>
<td>15.68 ounces</td>
<td>7.5 ounces</td>
</tr>
<tr>
<td>Prefer</td>
<td>✔</td>
<td>☐</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Feature</th>
<th>TG-4</th>
<th>G9</th>
</tr>
</thead>
<tbody>
<tr>
<td>Waterproof</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Prize</td>
<td>$249.99</td>
<td>$399.99</td>
</tr>
<tr>
<td>Weight</td>
<td>7.36 lb</td>
<td>7.5 lb</td>
</tr>
<tr>
<td>Prefer</td>
<td>☐</td>
<td>✔</td>
</tr>
</tbody>
</table>

We recommend:

• How do I pick the **next (1st) question** to obtain the largest reduction of uncertainty or “variance” on preferences

• Compensatory model estimation (**part-worths**), not just assortment

Experimental Design with MIP
Next Question To Reduce “Variance”: Bayesian

- Prior Distribution of $\beta$

<table>
<thead>
<tr>
<th>Feature</th>
<th>SX530</th>
<th>RX100</th>
</tr>
</thead>
<tbody>
<tr>
<td>Zoom</td>
<td>50x</td>
<td>3.6x</td>
</tr>
<tr>
<td>Prize</td>
<td>$249.99</td>
<td>$399.99</td>
</tr>
<tr>
<td>Weight</td>
<td>15.68 ounces</td>
<td>7.5 ounces</td>
</tr>
<tr>
<td>Prefer</td>
<td>✓</td>
<td>□</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Feature</th>
<th>TG-4</th>
<th>Galaxy 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Waterproof</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Prize</td>
<td>$249.99</td>
<td>$399.99</td>
</tr>
<tr>
<td>Viewfinder</td>
<td>Electronic</td>
<td>Optical</td>
</tr>
<tr>
<td>Prefer</td>
<td>✓</td>
<td>□</td>
</tr>
</tbody>
</table>

- Bayesian Update
- MCMC
- Posterior Distribution

- Black-box objective: Question Selection = Enumeration 😞
- Question selection by Mixed Integer Programming (MIP)
Paradoxes, Contradictions, and the Limits of Science

Many research results define boundaries of what cannot be known, predicted, or described. Classifying these limitations shows us the structure of science and reason.

Noson S. Yanofsky

“A computer would have to check all these possible routes to find the shortest one.”
MIP = Avoid Enumeration

- Number of tours for 49 cities = \( \frac{48!}{2} \approx 10^{60} \)
- Fastest supercomputer \( \approx 10^{17} \) flops
- Assuming one floating point operation per tour:
  \( > 10^{35} \) years \( \approx 10^{25} \) times the age of the universe!

- How long does it take on an iphone?
  - Less than a second!
  - 4 iterations of cutting plane method!
  - Dantzig, Fulkerson and Johnson 1954 did it by hand!
  - For more info see tutorial in ConcordeTSP app
  - **Cutting planes** are the key for effectively solving (even NP-hard) MIP problems in practice.
50+ Years of MIP = Significant Solver Speedups

• Algorithmic Improvements (Machine Independent):
  – Commercial, but free for academic use

• (Reasonably) effective free / open source solvers:
  – GLPK, **COIN-OR (CBC)** and SCIP (only for non-commercial)

• Easy to use, fast and versatile modeling languages
  – Julia based JuMP modelling language

• Linear MIP solvers very mature and effective:
  – Convex nonlinear MIP getting there (even MI-SDP!), quadratic nearly there
### Choice-based Conjoint Analysis (CBCA)

<table>
<thead>
<tr>
<th>Feature</th>
<th>Chewbacca</th>
<th>BB-8</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wookiee</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Droid</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Blaster</td>
<td>Yes</td>
<td>No</td>
</tr>
</tbody>
</table>

**I would buy toy**
- ✔️ Chewbacca
- □ BB-8

**Product Profile**

\[
\begin{pmatrix}
0 \\
1 \\
0
\end{pmatrix} = x^2
\]
MNL Preference Model

• Utilities for 2 products, n features (e.g. n = 12)

\[
U_1 = \beta \cdot x^1 + \epsilon_1 = \sum_{i=1}^{n} \beta_i x_i^1 + \epsilon_1
\]

\[
U_2 = \beta \cdot x^2 + \epsilon_2 = \sum_{i=1}^{n} \beta_i x_i^2 + \epsilon_2
\]

• Utility maximizing customer: \( x^1 \geq x^2 \) \( \iff \) \( U_1 \geq U_2 \)

• Noise can result in response error:

\[
L(\beta \mid x^1 \geq x^2) = P( x^1 \geq x^2 \mid \beta) = \frac{e^{\beta \cdot x^1}}{e^{\beta \cdot x^1} + e^{\beta \cdot x^2}}
\]
Next Question To Reduce “Variance”: Bayesian

Prior Distribution of $\beta$

<table>
<thead>
<tr>
<th>Feature</th>
<th>SX530</th>
<th>RX100</th>
</tr>
</thead>
<tbody>
<tr>
<td>Zoom</td>
<td>50x</td>
<td>3.6x</td>
</tr>
<tr>
<td>Prize</td>
<td>$249.99</td>
<td>$399.99</td>
</tr>
<tr>
<td>Weight</td>
<td>15.68 ounces</td>
<td>7.5 ounces</td>
</tr>
<tr>
<td>Prefer</td>
<td>✓</td>
<td>✗</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Feature</th>
<th>TG-4</th>
<th>Galaxy 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Waterproof</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Prize</td>
<td>$249.99</td>
<td>$399.99</td>
</tr>
<tr>
<td>Viewfinder</td>
<td>Electronic</td>
<td>Optical</td>
</tr>
<tr>
<td>Prefer</td>
<td>✓</td>
<td>✗</td>
</tr>
</tbody>
</table>

MNL Preference Model

$$\beta \cdot x^1 \geq \beta \cdot x^2$$

Logic Regression

$$\beta \cdot z \geq 0$$

$$z = x^1 - x^2$$

Experimental Design with MIP
“Linear” Experimental Design

\[ \text{Unknown } \beta \in \mathbb{R}^n \]

\[ \left\{ \beta \cdot z^i \right\}_{i=1}^q \]

Questions:

\[ Z = \begin{bmatrix} z^1 & \ldots & z^q \end{bmatrix}^T \in \mathbb{R}^{q \times n} \]

Model:  \[ P(Y \mid \beta, Z) = L(Y \mid \beta, Z) = \prod_{i=1}^q h(y^i, \beta \cdot z^i) \]

Objective: Choose \( Z \) to learn \( \beta \) “fast”
Bayesian Framework

Prior distribution

Answer likelihood

Posterior distribution

$\beta \sim N(\mu, \Sigma)$

$L(Y \mid \beta, Z)$

$g(\beta \mid Y, Z) = \frac{\phi(\beta \mid \mu, \Sigma) L(Y \mid \beta, Z)}{\int_{\mathbb{R}} \phi(\beta \mid \mu, \Sigma) L(Y \mid \beta, Z) \, d\beta}$

“fast” = minimize posterior “variance”
Goal = Minimize Expected Posterior Variance $f(Z)$

$$f(Z) = \mathbb{E}_Y \left\{ (\det \text{cov} (\beta | Y, Z))^{1/m} \right\}$$

Possible solution approaches:

$$g(\beta | Y, Z) \propto \phi(\beta ; \mu, \Sigma) L(Y | \beta, Z)$$

$$I(\beta | Y, Z) := -\frac{\partial^2}{\partial \beta \partial \beta} \ln g(\beta | Y, Z) \propto \sum^{-1} - \frac{\partial^2}{\partial \beta \partial \beta} \ln L(Y | \beta, Z)$$

$$\text{cov}(\beta | Y, Z) \approx I\left(\hat{\beta} | Y, Z\right)^{-1}, \quad \mathbb{E}_{\beta \sim N(\mu, \sigma)} \left\{ I(\beta | Y, Z)^{-1} \right\}$$

$$\max_Z \mathbb{E}_Y \left\{ \left( \det I\left(\hat{\beta} | Y, Z\right) \right)^{1/m} \right\}$$
A Really Good Case = Linear Regression

\[
f (Z) = \mathbb{E}_Y \left\{ (\text{det cov} (\beta \mid Y, Z))^{1/m} \right\}
\]

\[
y^i = \beta \cdot z^i + \epsilon_i, \quad \epsilon_i \sim N(0, 1)
\]

\[
g (\beta \mid Y, Z) = \phi (\beta ; \mu', \Sigma')
\]

\[
\Sigma' = \text{var} (\beta \mid Y, Z) = (Z^T Z + \Sigma^{-1})^{-1}
\]

\[
\min_Z f (Z) = \max_Z \left( \text{det} \left( Z^T Z + \Sigma^{-1} \right) \right)^{1/m}
\]

\[
Z \text{ discrete} \quad \rightarrow \quad \text{MISDP or MISOCP for } m = n
\]
A Relatively Good Case = Few Questions

\[ f(Z) = \mathbb{E}_Y \left\{ (\det \text{cov} (\beta \mid Y, Z))^{1/m} \right\} \]

\[ Z = \{ z^i \}_{i=1}^q \subseteq \mathbb{R}^n, \quad q \ll n \]

\[ \mathbb{E} (\beta \mid Y, Z) = m \left( Y, \{ \mu \cdot z^i \}_{i=1}^q, \left\{ z^iT \sum z^j \right\}_{i,j=1}^q \right) \]

\[ \text{cov} (\beta \mid Y, Z) = M \left( Y, \{ \mu \cdot z^i \}_{i=1}^q, \left\{ z^iT \sum z^j \right\}_{i,j=1}^q \right) \]

\[ f(Z) = f \left( \{ \mu \cdot z^i \}_{i=1}^q, \left\{ z^iT \sum z^j \right\}_{i,j=1}^q \right) \]
A Relatively Good Case = Few Questions

\[ f(Z) = E_Y \left\{ (\det \text{cov} (\beta | Y, Z))^{1/m} \right\} \]

\[ Z = f \left( \{ \mu \cdot z \}_{i=1}^q, \{ z \}_{i,j=1}^q \right) \]

Only requirements:

- \( \beta \sim N(\mu, \Sigma) \)
- \( L(Y | \beta, Z) = \prod_{i=1}^q h(y^i, \beta \cdot z^i) \)

✓ Logistic regression with small \( q \)

Experimental Design with MIP
Question Selection for CBCA

- "Variance" = D-Efficiency:
  \[ f(x^1, x^2) := \mathbb{E}_{\beta, x^1 \succeq \succeq x^2} \left( \det(\Sigma_i)^{1/p} \right) \]

- Non-convex function

- Without previous slide, even evaluation requires \( \dim(\beta) \) - dimensional integration

\[ \beta \sim N(\mu, \Sigma) \]

\[ x^1 \succeq x^2 \]

\[ \text{cov}(\beta) = \Sigma_1 \]

\[ \text{cov}(\beta) = \Sigma_2 \]
D-efficiency Simplification for CBCA

- D-efficiency = Non-convex function \( f(d, v) \) of
  - distance: \( d := \mu \cdot (x^1 - x^2) \)
  - variance: \( v := (x^1 - x^2)' \cdot \sum \cdot (x^1 - x^2) \)

Can evaluate \( f(d, v) \) with 1-dim integral 😊
Simplification = Trade-off for known criteria

\[(\beta - \mu)' \cdot \Sigma^{-1} \cdot (\beta - \mu) \leq r\]

• Choice balance:
  – Minimize distance to center
    \[\mu \cdot (x^1 - x^2)\]

• Postchoice symmetry:
  – Maximize variance of question
    \[(x^1 - x^2)' \cdot \Sigma \cdot (x^1 - x^2)\]
Optimization Model

\[ \min \quad f(d, \nu) \]

\[ \text{s.t.} \]

\[ \mu \cdot (x^1 - x^2) = d \] ✓

\[ (x^1 - x^2)' \cdot \sum \cdot (x^1 - x^2) = \nu \] ✗

\[ A^1 x^1 + A^2 x^2 \leq b \] ✓

linearize \( x^k_i \cdot x^l_j \)

\[ x^1 \neq x^2 \] ✓

\[ x^1, x^2 \in \{0, 1\}^n \]
Technique 2: Piecewise Linear Functions

- D-efficiency = Non-convex function \( f(d, v) \)
  - distance: \( d := \mu \cdot (x^1 - x^2) \)
  - variance: \( v := (x^1 - x^2)' \cdot \sum \cdot (x^1 - x^2) \)

Can evaluate \( f(d, v) \) with 1-dim integral 😊

Piecewise Linear Interpolation

MIP formulation
MIP-based Adaptive Questionnaires

<table>
<thead>
<tr>
<th>Feature</th>
<th>SX530</th>
<th>RX100</th>
</tr>
</thead>
<tbody>
<tr>
<td>Zoom</td>
<td>50x</td>
<td>3.6x</td>
</tr>
<tr>
<td>Prize</td>
<td>$249.99</td>
<td>$399.99</td>
</tr>
<tr>
<td>Weight</td>
<td>15.68 ounces</td>
<td>7.5 ounces</td>
</tr>
<tr>
<td>Prefer</td>
<td>✓</td>
<td>□</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Feature</th>
<th>TG-4</th>
<th>G9</th>
</tr>
</thead>
<tbody>
<tr>
<td>Waterproof</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Prize</td>
<td>$249.99</td>
<td>$399.99</td>
</tr>
<tr>
<td>Weight</td>
<td>7.36 lb</td>
<td>7.5 lb</td>
</tr>
<tr>
<td>Prefer</td>
<td>□</td>
<td>✓</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Feature</th>
<th>TG-4</th>
<th>Galaxy 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Waterproof</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Prize</td>
<td>$249.99</td>
<td>$399.99</td>
</tr>
<tr>
<td>Viewfinder</td>
<td>Electronic</td>
<td>Optical</td>
</tr>
<tr>
<td>Prefer</td>
<td>✓</td>
<td>□</td>
</tr>
</tbody>
</table>

\[
\mathbb{E}\left(\beta \mid Y, X^1, X^2\right) \\
\text{cov}\left(\beta \mid Y, X^1, X^2\right)
\]

- Optimal one-step look-ahead moment-matching approximate Bayesian approach.
Optimal One-Step Look-Ahead

Prior distribution

\[ \beta \sim N \left( \mu^i, \Sigma^i \right) \]

\[ \min_{x^1, x^2} f \left( x^1, x^2 \right) \]

- Solve with MIP formulation
Moment-Matching Approximate Bayesian Update

Answer likelihood

Prior distribution

Posterior distribution

\[ \beta \sim N \left( \mu^i, \Sigma^i \right) \]

- \[ \mu^{i+1} = \mathbb{E} \left( \beta \mid y, x^1, x^2 \right) \]
- \[ \Sigma^{i+1} = \text{cov} \left( \beta \mid y, x^1, x^2 \right) \]

\[ \beta^{\text{approx.}} \sim N \left( \mu^{i+1}, \Sigma^{i+1} \right) \]

1-dim integral
Computational Experiments

• 16 questions, 2 options, 12 and 24 features
• Simulate MNL responses with known $\beta^*$
• Question Selection
  – MIP-based using CPLEX and open source COIN-OR solver
  – Knapsack-based geometric Heuristic by Toubia et al.
• Time limits of 1 s and 10 s
• Metrics:
  – Estimator variance $= \left( \text{det} \, \text{cov} \left( \beta \mid Y, X^1, X^2 \right) \right)^{1/2}$
  – Estimator distance $= \left\| \mathbb{E} \left( \beta \mid Y, X^1, X^2 \right) - \beta^* \right\|_2$
  – Computed for true posterior with MCMC
Results for 12 Features, 1 s time limit

**Estimator Variance**

- **Heuristic** (Avg. = 0.04 s, Max = 0.61s)
- **COIN-OR** (Avg. = 0.93 s, Max = 1s)
- **CPLEX** (Avg. = 0.21 s, Max = 0.48s)
Does it Scale? Results for 24 features

**Estimator Variance**

- **Heuristic** (Avg. = 0.19 s, Max = 3s)
- **CPLEX 1s** (Avg. = 1 s, Max = 1s)
- **CPLEX 10s** (Avg. = 7.7 s, Max = 10s)
Some improvements for 24 features

**Estimator Variance**

- **Heuristic**: (Avg. = 0.19 s, Max = 3s)
- **CPLEX 1s**: (Avg. = 1 s, Max = 1s)
- **CPLEX 10s**: (Avg. = 7.7 s, Max = 10s)

**Estimator Distance**

**Experimental Design with MIP**
Summary and Main Messages

• Always choose Chewbacca!

• MIP can now “solve” challenging problems in practice
  – Even in near-real time
  – Appropriate domain expertise can be crucial for MIP’ing
  – Commercial solvers best, but free solvers reasonable
  – Integration into complex systems easy with JuMP
  – Some scalability: get the most out of “small” data

• Adaptive Choice-based Conjoint Analysis
  – Improves on existing geometric methods